Report of the October 2003 Meeting of the Science Archives Working Group

The SAWG held their fourth meeting on October 27-28, 2003 at NASA HQ, with the following members present: Julian Borrill, Joel Bregman (Chair), Roger Brissenden, Damian Christian, Eric Feigelson, Menas Kafatos, Bill Oegerle (Deputy Chair), Tom McGlynn, Sally Oey, and Rick White, along with the NASA HQ personnel Jeff Hayes, Alan Smale, and Joe Bredekamp.

Reports of the Archive Centers

HEASARC

Concerning other HEASARC activities, the SAWG endorses the development of links between its data holdings and the published literature through arrangements with the refereed journals and Astrophysics Data System. We also recommend that the HEASARC improve its efforts to inform the science community of its newer Web service products, such as the Virtual Observatory Data Inventory Service and the Hera data analysis environment.

LAMBDA and HEALPIX

The SAWG received a brief update on some outstanding issues concerning the Legacy Archive for Microwave Background Data Analysis. Our major concern continues to be connection between LAMBDA and the HEALPix software package. HEALPix has become the standard pixelization for CMB maps, including those from NASA's current WMAP and future Planck satellite missions. The associated software is fundamental to handling CMB data, but its long-term development and maintenance is currently unsupported, and proceeding on a best-effort basis only. Since the HEALPix team leader has recently moved to JPL, the SAWG sees this as an ideal opportunity to secure the future of HEALPix by providing long-term support from NASA. In addition it was noted that the recommended LAMBDA user group has still to be convened.

MAST

Marc Postman presented an update on MAST activities. MAST has ingested the first GALEX data release, and is working with the GALEX team to develop GMAX (GALEX Multi-Archive Cross-Correlation Tool). Data from CHIPS is also soon anticipated, although the volume is low. The WFPC2 associations are now co-added and cleaned of cosmic rays.

There is debate on prioritizing the MAST query capability to support object class queries as suggested by users, as compared to strictly observable parameters. The SAWG favors adding only the broadest categories of object class. Prior to our meeting, data retrieval had been slower than desired, in spite of a June '03 RAID system installation, although new HST/FUSE processing software will be installed in December 2003. Improvements during the September-December 2003 are expected to speed up the system considerably and there is good evidence of that already. The SAWG suggests a review of future retrieval needs and load scalability. We also note that MAST has no budgeted E/PO component, in contrast to HEASARC and IRSA; we suggest maintaining cognizance of this issue.

Software Development Issues

Presentations relating to the HEASARC, LAMBDA, and MO&DA pointed out some issues relating to software development. A few software tools, developed under the AISRP program, have become extremely widely used and for many investigators, have become powerful and integral elements of the astronomer's toolbox when investigating archival data. Notable among these are the ds9 FITS interactive image display tool and the SkyView multiwavelength visualization Web service (although there are other examples). The development of these packages took a number of years (under AISRP), but the SAWG felt that they had become important enough that there should be a more permanent commitment to their future maintenance and development. This responsibility could be assigned to a particular archive center, provided that funding and expertise were available and appropriate.

These two examples raise the general issue of how and when an experimental software package merits more permanent attention and where it can be accessed during development. This last issue is the "visibility" of the software, in that if it is offered on an archive web page, it is highly visible, but if it resides on someone's home page, it will be known to a much smaller group of people. A possible approach to dealing with this is to have a web page at each archive center devoted to experimental software whereby it would merely provide the links to the software, which may exist at personal sites. This would be a low-budget approach to dealing with the problem of providing a "home" for the software. The archive centers could keep track of the number of unique users in evaluating whether a piece of software should be deleted from the page (or promoted). Such actions would require some discussion between the AISRP program and the ADEC.

The ADEC, and Virtual Observatory Activities

Nick White presented a status of the Astrophysics Data Centers Executive Committee (ADEC) and emphasized the importance of maintaining the momentum associated with the Celestial Navigator proposal - a proposal for a NASA virtual observatory precursor interoperability initiative. We continue to endorse ADEC as a means of coordinating NASA Space Science Data Center efforts and as way of leveraging the invested resources, particularly in the area of interoperability. We strongly support Code S in their efforts to provide initial funding for this initiative starting in FY04. We also support encouraging the community to propose through ASIRP for the development of interoperability tools.

Given the importance of the role of the Data Centers in establishing a NASA interoperability initiative, steps should be taken to ensure that these plans and their context are communicated to the Senior Review. We support providing the Data Centers with guidelines for a common proposal structure that includes a request for a discussion of their participation in ADEC and their interoperability activities and plans. We encourage the ADEC to coordinate this aspect (at least) of the Data Center proposals to the Senior Review. We also encourage NASA to provide a statement to the Senior Review describing the role of the ADEC and NASA's plans for interoperability initiatives.

SWIFT Project Data Management Plan

The SAWG appreciated the comprehensive discussion of the Swift ground systems presented by Frank Marshall. In discussing the Swift PDMP, the only substantive issue was the level of detail with which the PDMP should describe the data flow and archiving plans. Several

members of the SAWG felt that a basic understanding of the data flows was not manifest in the document, and was actually somewhat obscured by the detailed discussion in the PDMP. The SAWG recognizes that instructions for preparing PDMP's are under review and strongly encourages this review to include specific guidelines for the level of detail to which the PDMP should go.

The SAWG echoes Dr. Marshall's concern that development of test data sets early in the mission is extremely desirable, while recognizing that the very tight timeline for the Swift mission may have made this difficult.

NASA Herschel Science Center

In a previous meeting, the SAWG expressed some concern over what appeared to be high staffing levels at the NASA Herschel Science Center (NHSC). We thank George Helou for providing a thorough description of the funding at the NHSC. The committee is considerably more comfortable with the planned levels based on Helou's presentation. In the absence of an MOU with ESA defining the NHSC tasks, however, it is not possible for us to assess whether a smaller HSC effort balanced by a larger ESA effort would accomplish the same goals. Once an MOU with ESA is in place (soon, we hope), it will be necessary to review the planned work at the NHSC and its relation to work at ESA to ensure that there is minimal duplication of effort.

On other matters, the SAWG was disappointed that the PDMP from GALEX has not be received, despite repeated requests, and we hope that this is corrected soon.

The next meeting of the SAWG is expected to take place in May 2004 and we welcome suggestions from the SEUS and the OS for future topics to be addressed.